Parts of this chapter were published as part of a review:

J.S. wrote the review.
The pneumococcus – a commensal gone rogue

*Streptococcus pneumoniae* (the pneumococcus) is a Gram-positive human commensal that colonizes the nasopharynx. There, the pneumococcus is mostly found in complex biofilms, formed together with several other members of the nasopharyngeal microbiome [1,2]. Pneumococcal carriage rates peak at 2–3 years of age, when close to 60% of children are colonized by the pneumococcus at any one point [3]. Even more tellingly, Gray et al. showed that, by the age of 2, more than 95% of children had been colonized at least once [4]. Afterwards, although dependent on demographic, temporal and social factors [5], the average carriage rate drops to below 10% in adults [5,6]. Fortunately, in the nasopharynx, the pneumococcus is predominantly harmless. However, infrequently, it can leave its preferred niche and invade other parts of the human body, including the lungs, cerebrospinal fluid and blood. There, it poses a serious threat as it can cause potentially lethal diseases, such as pneumonia (lungs), meningitis (cerebrospinal fluid) or sepsis (blood) [6]. As a result, the pneumococcus is responsible for more than a million deaths every year, especially among children, the elderly and immunocompromised individuals [7]. Combined with the upsurge of antimicrobial resistant pneumococcal strains, this has been reason for the World Health Organization, in 2017, to place *Streptococcus pneumoniae* among the 12 ‘priority pathogens’ for future research ([http://www.who.int/medicines/publications/WHO-PPL-Short_Summary_25Feb-ET_NM_WHO.pdf](http://www.who.int/medicines/publications/WHO-PPL-Short_Summary_25Feb-ET_NM_WHO.pdf)).

Pneumococcal virulence factors

To be able to effectively colonize the human nasopharynx, the pneumococcus depends on a wide variety of compounds, collectively referred to as virulence factors. Not surprisingly, many of these virulence factors have also been implicated in disease [8]. Extensive reviews on pneumococcal virulence factors are available [8–10] and just a few are highlighted here. Firstly, the pneumococcus expresses a polysaccharide capsule that both facilitates access to human epithelial cells [11] and protects the bacterial cell from phagocytosis by the host [12]. Pneumococci with a wide variety of capsular (*cps*) genes, leading to significant differences on the molecular level, have been found and are referred to as serotypes. To date, more than 90 serotypes are known, with new ones still being discovered [13].

A second important virulence factor is pneumolysin, which can actively form pores in the membranes of diverse host cell types [14]. The release of the cytoplasmically expressed pneumolysin from pneumococcal cells is primarily enabled by LytA-mediated autolysis [15]. Other virulence factors include PavB [16] and CbpA (PspC) [17], involved in adherence, and PspA, which protects the pneumococcus against lactoferrin-mediated killing [18]. Finally, the more recently identified pneumococcal histidine triad proteins (PhtA, PhtB, PhtD and
PhtE) were shown to inhibit complement deposition on the pneumococcal cell surface [19], adding another layer of protection against the human immune system.

**Prevention and treatment of pneumococcal infections**

The most effective approach to lowering mortality rates as a result of pneumococcal infections is to prevent such infections in the first place. To that end, several generations of pneumococcal vaccines have been developed since the 1940s. The first vaccine family to be developed, the pneumococcal polysaccharide vaccines (PPSVs), is based on purified capsular polysaccharides [20]. The most recent PPSV, called Pneumovax® 23, grants protection against 23 serotypes, making up 85–90% of all serotypes causing invasive disease in the United States, as estimated in 1999 [21]. Although PPSVs have proven effective in preventing invasive disease, they were shown to have limited efficacy especially in the elderly and children under 2 years of age [22]. For that reason, a second generation of vaccines was introduced: pneumococcal conjugate vaccines (PCVs), in which pneumococcal polysaccharides are attached to a carrier protein. In contrast with PPSVs, which represent T-cell-independent antigens [21], PCVs incite a T-cell-dependent response, enhanced by the production of memory B cells. As a result, invasive pneumococcal disease was significantly reduced. In children under 2 years of age, a striking drop in disease rate of 69% was observed [23]. An additional advantage of PCVs over PPSVs is that, besides reduced disease rates, also carriage rates are reduced by PCVs, leading to the indirect protection of non-vaccinated individuals [24].

A disadvantage of the conjugate vaccines and, to a lesser extent, PPSVs lies in the practical limitation of the number of serotypes that can be included. The introduction of the first PCV (PCV7), covering 7 serotypes, resulted in a phenomenon called serotype replacement [25]: the removal of vaccine-included serotypes opened a niche that could be filled by non-vaccine serotypes (or competing commensals). Therefore, PCV13 was introduced to also include the newly dominant serotypes [26]. Although proven effective, it remains to be seen whether a new wave of serotype replacement will reduce the long-term efficacy of PCV13.

To address the issue of serotype replacement, there is a need for antigens that are more conserved among pneumococci, the most promising candidates being surface-exposed or secreted pneumococcal proteins. Candidate proteins under study include the previously mentioned virulence factors pneumolysin [27], PspA [28], CbpA (PspC) [29], and pneumococcal histidine triad protein PhtD [30,31].

Finally, pneumococcal infections can be treated with a variety of antibiotics. Most commonly, macrolides or beta-lactams are prescribed (e.g. a
cocktail of the beta-lactam amoxicillin and beta-lactamase inhibitor clavulanic acid). Alternatively, members of the beta-lactam subfamily of cephalosporins are used. However, in specific cases, other classes of antibiotics are administered, including fluoroquinolones (e.g. levofloxacin). It is relevant to note, with regard to Chapters 5 and 6, that S. pneumoniae may also encounter antibiotics that are not very effective in killing it. An example of this is the beta-lactam aztreonam, that is mostly effective against Gram-negative bacteria, such as Pseudomonas aeruginosa.

**Pneumococcus – the escape artist**

Despite the effectiveness of both antibiotic therapies and vaccination programs, there is reason for alarm. The pneumococcus has proven to be a formidable adversary and, like many other members of the human microbiome, hard to completely eradicate. One of the reasons why the pneumococcus is so persistent, is its extraordinary genomic plasticity, which has been shown to facilitate the evasion of the immune system. Specifically, in addition to serotype replacement discussed above, pneumococci are also able to actively change their capsular type through recombination [32]. This further compromises the expected sustainability of vaccines based on capsular polysaccharides. Moving the focus to more conserved antigens, however, Croucher et al. showed that several surface-exposed proteins, including PspA and CbpA (PspC), displayed accelerated rates of evolution [33,34]. Additionally, in Chapter 2, we show that histidine triad proteins are also subject to recombination-mediated sequence variation. Together, these observations suggest that caution should be taken to also monitor the long-term effect of new vaccines on pneumococcal antigen allele frequencies. To add insult to injury, the increasing level of multidrug resistant pneumococci is frightening. Illustratively, in a 2009 study from Japan, Imai et al. showed that 91% of carriage serotypes and 53% of medium carriage and invasive serotypes were resistant to three or more antibiotics [35].

**Pneumococcal competence for genetic transformation**

Undoubtedly, a major contributing factor to this Houdini act [36] is the ability of the pneumococcus to become naturally competent for genetic transformation, allowing them to take up exogenous DNA and integrate it into their own genome (i.e. transformation). While bacterial competence is traditionally specifically referring to the transformation process, the pneumococcal competent state has been reported to encompass many other functionalities, including DNA repair, bacteriocin production and heat shock response [37,38]. This diversity of activated functions is relevant in light of the fact that a broad spectrum of antimicrobial compounds (causing various forms of stress) can actually induce competence development (Chapters 5, 6; [39]). Since the pneumococcus lacks
several other reported prokaryotic stress responses, such as the widely conserved SOS response [40], the competent state was speculated to serve as an important general stress response mechanism [41,42].

The importance of gene copy numbers

How genome organization and gene function are connected

For decades, the importance of genome organization has been recognized. Virtually every process that interacts directly or indirectly with the chromosome has left its marks during the course of genome evolution. It has become clear that the order and orientation of features on a chromosome as well as the three-dimensional structure of the chromosome is of importance to a cell. Numerous examples of the interplay between genome organization and cellular processes are available. For example, essential genes tend to be located on the strand that is transcribed in the same direction as in which replication proceeds [43].

However, the importance of the genomic location of key elements is still often underestimated. In fact, very little attention is given to the many different ways in which genomic location can impact cell biology. In a review, we provided an extensive overview of the various mechanisms by which the exact genomic location of a feature can play a role in the regulatory landscape and development of bacterial cells [44]. More specifically, we focused on processes in which gene copy number or, more accurately, genome-wide copy number distributions play a role. It is a well-established fact in eukaryotes that having an abnormal number of chromosomes (aneuploidy), leading to atypical gene copy numbers, can have detrimental effects, a well-known example being Down syndrome (trisomy 21 in humans) [45]. Additionally, the need for female mammals to silence one of their two copies of the X-chromosome, underlines the importance of DNA copy numbers [46,47]. Furthermore, amplification of specific nutrient transporter genes in *Saccharomyces cerevisiae* was observed to enhance fitness in nitrogen–limited conditions [48]. The correlation between copy number and gene expression implied by these examples was confirmed recently by Chen and Zhang, who showed that the timing of replication of a gene influences its final expression level in yeast [49]. Nevertheless, copy number effects are still only rarely considered in prokaryotes. During bacterial cell cycle progression, copy numbers around the chromosome fluctuate periodically. Both the periodicity [50] and the amplitude [51,52] of this fluctuation can be employed to regulate certain processes in the cell. Furthermore, global or local (e.g. compartmentalization) distortions of copy number fluctuations can be involved in bacterial ‘decision-making’ and even play an important role during virulence [52].
Replication–associated copy number fluctuations

The majority of bacteria have their DNA organized on a single, circular chromosome, replication of which starts at a well-defined origin of replication (oriC). From there, replication proceeds symmetrically in both directions around the chromosome and is terminated at the opposite end (the ter region) of the molecule, where both replication machineries (forks) meet. As a result, the various genes and other features on the chromosome are replicated in a fixed order, leading to periodic fluctuations of their copy numbers that are repeated every cell-cycle. After termination of replication, cells still need a specific amount of time to finish cell division (the D-period; [51]). The initiation of new rounds of replication is tightly regulated by a variety of factors [53–56]; this ensures there is exactly one initiation event each cell cycle, timed in such a way that replication and cell division are properly coordinated. When growth is sufficiently slow, cells have enough time to start and finish DNA replication within one cycle and local copy numbers will generally only fluctuate between one and two copies of a certain region (Figure 1A). Some bacteria, however, have the capacity to grow so fast that replication of their entire chromosome cannot be executed within one cell cycle [57]. In this case, cells engage in multifork replication; before a replication fork has finished, a new replication initiation event takes place (still exactly once per cell cycle) at all (≥ 2) copies of oriC simultaneously, resulting in copy numbers of oriC–proximal regions of more than 2 (Figure 1B). For example, fast-growing Escherichia coli cells have been observed to contain up to 8 origins [58]. Since there is a clear correlation between gene copy number and gene expression [59–61], these fluctuations are relevant to a cell’s transcriptome as is exemplified by the various cases discussed below and in our review [44].

Function–associated gene order

The amplitude of a gene’s copy number fluctuation will thus depend both on its genomic location, relative to oriC, and on growth rate. The impact of these dependencies is illustrated by the fact that translocations and chromosomal inversions preferentially occur in a copy-number–neutral fashion (i.e. symmetrical with respect to oriC or ter) [62–64], as also observed in Chapter 2. Another example of the importance of gene order is the strong conservation of the oriC–proximal colocalization of important growth factors involved in replication, transcription and translation [57,65,66]. The colocalization of these factors can be explained by a combination of the importance of their stoichiometry on the one hand and functional compartmentalization on the other. However, the fact that they are virtually always found close to the origin of replication rather reflects the cells’ need to correlate their expression with their requirement; when growth conditions improve, cells may switch to multifork replication, automatically boosting the expression of these essential growth factors due to the resulting
Figure 1. Replication-associated gene copy numbers. Simulated gene copy number distributions throughout the cell cycle (A and B). Each arm of the chromosome has been divided into four quartiles, which are color-coded based on their oriC-proximity. The height of each colored area in the graphs represents the average copy number within the corresponding quartile; as the replisome moves through a quartile, the corresponding graph area steadily increases in height until it is exactly doubled (i.e. the entire quartile is replicated), while the other areas maintain their height. The areas describing the copy number development of the four quartiles are stacked, so their combined height reflects the total DNA content of a cell. Average copy numbers of each quartile at 10%, 50% and 90% of the cell cycle are shown in the plots. The script to run the simulations is available upon request. Replication initiation is indicated by black arrows. (A) During relatively slow growth (replication time/cell cycle = 0.5, D-period = 10% of cell cycle), only one replication fork is present at a time on each arm of the chromosome (top) and gene copy numbers will fluctuate between 1 and 2 (bottom). (B) During relatively fast growth (replication time/cell cycle = 1.6, D-period = 10% of cell cycle), multifork replication occurs (top) and gene copy numbers can exceed 2 (bottom). (C) The oriC-proximal location of the *Vibrio cholerae* S10 ribosomal protein operon is important for fitness [52]. Top: translocation of these genes...
dosage increase. Recent work by Soler-Bistué et al. demonstrates the relevance of the genomic position of ribosomal protein genes on the large chromosome of the human pathogen *Vibrio cholerae*, which harbors two circular chromosomes (**Figure 1C**; [52]). They showed that translocation of a locus bearing half of all ribosomal protein genes from *oriC*-proximal to various sites further away from the origin of replication results in significant defects in growth and host-invasion capacity. It is worth noting that these defects specifically occur during relatively fast growth, where the difference in copy number between *oriC* and *ter*, and therefore the relative effect of translocation of the ribosomal protein genes, is the largest. Both defects are relieved when, instead of one, two copies of the locus are present at an *oriC*-distal site, effectively restoring absolute ribosomal gene copy numbers and consequently ribosome production levels. The fact that these genes are then no longer colocated with other important growth factors is, apparently, of lesser importance in this context.

Similarly, Sobetzko et al. demonstrated that nucleoid-associated proteins (NAPs) employed during exponential growth, together with their binding sites, show a tendency to be located closer to *oriC* than NAPs that act in (near-)stationary phase [66]. Simultaneously, they showed that genes with related functions have a propensity to be distributed at equal distances from *oriC*, without the necessity of being on the same arm of the chromosome [66]. Taken together, these observations underline that the variation in growth conditions encountered throughout evolution is directly reflected by the relative positioning on the chromosome of genes with related functions.

**Distortion of natural gene dosage fluctuation induces bacterial competence**

Whether or not a bacterium will perform multifork replication largely depends on the combination of its growth rate and its genome size. As discussed earlier, the *oriC*-proximal location of genes encoding important growth factors automatically correlates their production and requirement levels. A different way in which *oriC*-proximity is utilized is found in the pneumococcus (**Figure 2, Chapter 5**). With its relatively small genome (~2 Mb, **Chapter 2**), multifork replication in rapidly dividing cells has not been observed (**Chapter 5**). This situation changes, however, when replication fork progression is directly or indirectly perturbed
Chapter 1

General introduction

and slowed down. Since, as far as we know, there is no instantaneous feedback to the pneumococcal replication initiation system, new replication complexes may be loaded onto the genome before the stalled or slowed replication forks have finished, leading to increased dosage of oriC-proximal genes. Various factors can lead to this form of overinitiation: DNA damage (e.g. double-strand breaks induced by mitomycin C); insufficient functioning of type II topoisomerases, which are responsible for the relaxation of DNA required for replication forks to progress (e.g. induced by fluoroquinolone antibiotics); or limited nucleotide availability (e.g. induced by trimethoprim and hydroxyurea). *S. pneumoniae* makes use of this exceptional situation to activate competence (Chapter 5), allowing cells to take up and internalize exogenous DNA [41]. The activation of this system encompasses the expression of over a hundred genes (Chapter 4; [37,67,68]), blocks cell division [69], and thus represents a significant burden for the cell. It is therefore important for the cell to somehow regulate the activation of this system (also see Chapter 7). Despite the large number of genes eventually being activated, the on/off switch of competence is constituted by a positive-feedback loop containing a set of only five genes organized into two operons [70], *comAB* and *comCDE*. Very low-level basal expression occurs for both operons. ComC is a

Figure 2. Competence activation in *Streptococcus pneumoniae* due to dosage upshift of oriC-proximal regulator genes. The oriC-proximal location of early competence genes allows the pneumococcus to activate this state in response to replication stress (Chapter 5). Simulated development of copy number distribution during replication stress is shown in the bottom graph (bottom panel; same plotting parameters and (initially) same simulation parameters as in Figure 1A). Halfway the second cell cycle, replication stress is applied (red star; new replication rate is one-third of original replication rate), while timing of replication initiation events is unaltered (black arrows). Note that time units indicated with an asterisk are multiples of the cell cycle time in the absence of replication stress. Due to the oriC-proximal location of *comAB* and *comCDE*, their expression levels increase (bottom graph, top panel) and once a certain threshold activity is reached, competence is activated via the positive feedback loop in its regulatory system (top right).
41 residue peptide containing a double-glycine leader of 24 amino acids in length. Membrane-associated transporter complex ComAB exports ComC, cleaving off the leader peptide, and extracellularly releasing the 17 residue competence-stimulating peptide (CSP), which acts as a quorum-sensing autoinducer [71]. ComDE constitutes a typical two-component regulatory system; the membrane-bound histidine kinase ComD binds the extracellular CSP and consecutively transfers a phosphate group to the response regulator ComE, resulting in ComE~P. ComE~P then completes the positive feedback loop by enhancing expression of both comAB and comCDE [72]. Additionally, it induces the expression of comX, coding for the competence-specific sigma factor σX, required for the activation of the entire competence regulon (Chapter 4; [73]). However, processes like mRNA and protein degradation and dilution by growth will counteract this positive feedback loop and may prevent competence from switching on. Additionally, the autocatalytic efficiency of the system is dependent on medium parameters like pH. Only when the local extracellular CSP concentration exceeds a certain threshold, the positive feedback may outcompete the counteracting forces and competence gene expression may dramatically increase (possibly with several orders of magnitude). Hence, whether or not competence is activated depends on a complex set of parameters, including the copy numbers of comAB and comCDE; because of their oriC-proximal location on the chromosome (8° and -1°, respectively), relative overinitiation (e.g. due to replication fork stalling) can push up the dosage of early competence genes. We show, in Chapter 5, that even a slight increase in dosage, of below twofold, can suffice to reach threshold CSP concentrations and lead to competence activation. Interestingly, it was recently shown that the production of pneumococcal bacteriocins (pneumocins) is also potentiated by competence activation [74,75]. Since pneumocins play an important role in intra- and interspecies competition in their natural niche (the human nasopharynx), the gene-dosage-induced activation of competence may cause the composition of the nasopharyngeal flora to change, for better or for worse.

**Thesis outline**

**The genome and transcriptome of S. pneumoniae D39V**

Many genome assemblies of various pneumococcal strains are publicly available, including that of D39W [76], a direct derivative of strain NCTC 7466 (which can be obtained from the National Collection of Type Cultures of Public Health England). Although the strain used in our laboratory (D39V) is also derived from NCTC 7466, some differences were noticed between D39W and D39V. Because, in recent decades, DNA sequencing technology has made several leaps forward, we decided to perform de novo assembly of the S. pneumoniae D39V genome, using
Single-Molecule Real-Time (PacBio) sequencing. We annotated the resulting genome in high detail, including transcript boundaries, transcription regulatory elements and novel non-coding RNAs. The thereby assembled information was made available in a user-friendly genome browser, called PneumoBrowse, and is presented in Chapter 2.

Mapping Illumina RNA-seq data on the up-to-date genome annotation of strain D39V, we created a compendium of the pneumococcal transcriptome in 22 infection-relevant conditions. This rich data set of expression levels was also made available to the public, as PneumoExpress, and is discussed in Chapter 3. PneumoExpress includes a co-expression matrix, which reports on correlated gene expression throughout the studied conditions. As a proof of principle, we used the co-expression matrix to identify a novel competence-regulated gene, briC.

Pneumococcal competence

A subset of the PneumoExpress data set – a control sample and cells 3, 10 and 20 minutes after competence activation – was used to refine the competence regulon, which was previously determined using DNA microarray technology [37,38]. Using transcript boundaries (Chapter 2) and co-expression data (Chapter 3), we could largely attribute observed transcriptomic changes to specific affected promoters and thereby create a completer and more nuanced overview of differential expression during competence (Chapter 4).

As mentioned before, a variety of different antimicrobial compounds are known to promote pneumococcal competence development. However, only competence induction by aminoglycosides was understood mechanistically [77,78]. In this thesis, two other mechanisms for antibiotic-induced competence activation are presented. Firstly, compounds targeting DNA replication, such as fluoroquinolones, lead to shifted gene dosage distributions (as briefly discussed above) and increased expression of genes involved in competence regulation (Chapter 5). Secondly, beta-lactam aztreonam and beta-lactamase inhibitor clavulanic acid give rise to a chain-forming phenotype. We show in Chapter 6 that such a phenotype transforms competence regulation from global to local quorum-sensing, with reduced communication between different chains of cells. As a result, competence is promoted, albeit in a less synchronized fashion compared to untreated cells.
References
Singleton, R.J. et al. (2007) Invasive pneumococcal disease caused by nonvaccine serotypes among Alaska native children with high levels of 7-valent pneumococcal conjugate vaccine coverage. *JAMA* 297, 1784–1792


Alexander, J.E. et al. (1994) Immunization of mice with pneumolysin toxoid confers a significant degree of protection against at least nine serotypes of *Streptococcus pneumoniae*. *Infect. Immun.* 62, 5683–5688


Mackiewicz, P. et al. (2001) Flip–flop around the origin and terminus of replication in prokaryotic genomes. *Genome Biol.* 2:interactions, 1004.1–1004.4


Kjos, M. et al. (2016) Expression of *Streptococcus pneumoniae* bacteriocins is induced by antibiotics via regulatory interplay with the competence system. *PLoS Pathog.* 12, e1005422

76 Lanie, J.A. et al. (2007) Genome sequence of Avery’s virulent serotype 2 strain D39 of Streptococcus pneumoniae and comparison with that of unencapsulated laboratory strain R6. J. Bacteriol. 189, 38–51

77 Stevens, K.E. et al. (2011) Competence in Streptococcus pneumoniae is regulated by the rate of ribosomal decoding errors. mBio 2,
